73c

UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/786,475	02/24/2004	Karel Konvicka	100/1046-20	9333

31662   7590   10/26/2006

PERLEGEN SCIENCES, INC.
LEGAL DEPARTMENT
2021 STIERLIN COURT
MOUNTAIN VIEW, CA 94043

EXAMINER
NEGIN, RUSSELL SCOTT

ART UNIT	PAPER NUMBER
1631	

DATE MAILED: 10/26/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

	Application No.	Applicant(s)
Office Action Summary	10/786,475	KONVICKA, KAREL
	Examiner	Art Unit	
	Russell S. Negin	1631	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *15 August 2006*.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-79* is/are pending in the application.

    4a) Of the above claim(s) *14,15,19,78 and 79* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-13,16-18 and 20-77* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____ 6/14/04? 1/15/05? 7/17/06

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____ .

DETAILED ACTION

Election/Restrictions

Applicant's election of Group I in the reply filed on 15 August 2006 is

acknowledged. Because applicant did not distinctly and specifically point out the

supposed errors in the restriction requirement, the election has been treated as an

election without traverse (MPEP § 818.03(a)).

Claims 14, 15, 19, and 78-79 are withdrawn from further consideration pursuant

to 37 CFR 1.142(b) as being drawn to a nonelected Group or Species, there being no

allowable generic or linking claim. Election was made **without** traverse in the reply filed

on 15 August 2006.

Claims examined in this Office action are 1-13, 16-18, and 20-77.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of
matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the
conditions and requirements of this title.

Claims 1-13, 16-18, and 20-77 are rejected under 35 U.S.C. 101 because the

claimed invention is directed to non-statutory subject matter.

Upon consideration of the recent Official Gazette notice of November 22, 2005,

entitled, "Interim Guidelines for examination of patent applications for patent subject

matter eligibility," (www.uspto.gov/web/offices/com/sol/og/2005/week47/patgupa.htm),

the decision of the Office is to enact a 35 U.S.C. 101 rejection.

In regards to claims 1-13, 16-18, and 20-77, the instant claims are drawn to a genetic algorithm. A genetic algorithm is non-statutory unless the claims include a step of physical transformation, or if the claims include a useful, tangible and concrete result. It is important to note, that the claims themselves must include a physical transformation step or a useful, tangible and concrete result in order for the claimed invention to be statutory. It is not sufficient that a physical transformation step or a useful, tangible, and concrete result be asserted in the specification for the claims to be statutory. In the instant claims, there is no step of physical transformation, thus the Examiner must determine if the instant claims include a useful, tangible, and concrete result.

In determining if the instant claims are useful, tangible, and concrete, the Examiner must determine each standard individually. For a claim to be "useful," the claim must produce a result that is specific, substantial, and credible. For a claim to be "tangible," the claim must set forth a practical application of the invention that produces a real-world result. For a claim to be "concrete," the process must have a result that can be substantially repeatable or the process must substantially produce the same result again. Furthermore, the claim must recite a useful, tangible, and concrete result in the claim itself, and the claim must be limited only to statutory embodiments. Thus, if the claim is broader than the statutory embodiments of the claim, the Examiner must reject the claim as non-statutory.

Claims 1-13, 16-18, and 20-77 do not produce a tangible result. A tangible result requires that the claim must set forth a practical application to produce a real-world result. This rejection could be overcome by amendment of the claims to recite that a

result of the method is outputted to a display or a memory or another computer on a network, or by including a physical transformation.

As stated in the Official Gazette notice, "The tangible requirement does not necessarily mean that a claim must either be tied to a particular machine or apparatus or must operate to change articles or materials to a different state or thing. However, the tangible requirement does require that the claim must recite more than a Sec. 101 judicial exception, in that the process claim must set forth a practical application of that Sec. 101 judicial exception to produce a real-world result. Benson, 409 U.S. at 71-72, 175 USPQ at 676-77 (invention ineligible because had "no substantial practical application."). "[A]n application of a law of nature or mathematical formula to a . . . process may well be deserving of patent protection." Diehr, 450 U.S. at 187, 209 USPQ at 8 (emphasis added); see also Corning, 56 U.S. (15 How.) at 268, 14 L.Ed. 683 ("It is for the discovery or invention of some practical method or means of producing a beneficial result or effect, that a patent is granted . . ."). In other words, the opposite meaning of "tangible" is "abstract.""

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 21, 24, and 68 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claims 21, 24, and 68 recite the limitation, "the distance based clustering process

yields a minimum maximum standard deviation for a distribution..." in which the it is

unclear and ambiguous as to the metes and bounds of the phrase "minimum maximum"

signifies. Applicant is required to succinctly describe what is meant by this phrase.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-3, 5-7, 63, and 73 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Anderson et al. [Clinical Chemistry, volume 30, 1984, pages 2031-

2036] in view of Geever et al. (Proceedings of the National Academy of Science of the

USA, vol. 78, pp. 5081-5085, 1981).

1. A method for determining a genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: assigning the measure of the amount of the allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group, wherein the individual is determined to have the genotype assigned to the group.

2. A method as claimed in claim 1, wherein the method is computer-implemented.

3. A method as claimed in claim 1, wherein the genetic marker is a SNP position.

5. A method as claimed in claim 1, wherein the individual is a diploid organism.

6. A method as claimed in claim 5, wherein the diploid organism is a mammal.

7. A method as claimed in claim 6, wherein the mammal is a human.

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the genotype assigned to the group.

73. A computer readable medium comprising computer readable code for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of an allele of the genetic marker in the individual, and for carrying out the processes of: assigning the measure of the amount of an allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group and determining the individual to have the genotype assigned to the group.

Anderson et al teaches a computer-implemented method for probability clustering processes. As is stated on the last paragraph in column 2 of page 2032, "Two dimensional autoradiograms were scanned with an Optronics P-1000 microdensitometer at 100-um resolution, and analyzed with the Argonne TYCHO II system... This system (based on a VAX 11/780 computer)... applies film corrections, removes background, detects spots, and fits them with two dimensional Gaussian forms." The title of Anderson et al states, "Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis." The study indicates the presence of certain markers by the expression of relevant types of proteins. As is stated, in column 2 of page 2032, lines 12-19, "In this paper we discuss methods for generating statistical data sets in which the abundances of

numerous proteins are measured across samples (gels). Using actual 2-D gel data, we

have applied the techniques of principal component and cluster analysis to the problem

of determining relationships between a series of human cell types, and have made

some progress in demonstrating the potential of this approach." Figures 2 and 3 show

such probabilistic cluster analyses. A plurality of samples from a plurality of groups of

cells is analyzed.

Anderson et al. fails to show SNPs in general or explicitly mention the term

genotype. Additionally, Anderson et al. fails to conduct experiments on humans.

The objective of Geever et al. is that "a direct analysis of the sickle cell anemia

[in prenatal humans] should be possible by use of a restriction enzyme whose

recognition sequence is created or eliminated by the sickle cell mutation." (bottom of

page 5081, left column) Thus, this study finds the genotype of one individual (in this

case, the genotype related to sickle cell anemia (a SNP) in a prenatal human), from a

measure of the amount of a genetic marker (in this case the mutation site in the

hemoglobin gene shown in Table 1 on page 5082), designated as a zero amount or a

non-zero amount.

Figures 2, 3, and 4 of Geever show three different methods of determining

whether the genotype corresponding to the existence of a genetic marker indicative of

sickle cell anemia exists. Each of the respective measures is assigned an allele based

on the presence of bands in the autoradiograms. Genotypes could then be assigned

based on the presence or absence of these pluralities of markers. The plurality of

groups consists of samples for individuals that are either positive or negative for sickle

cell anemia. If relevant bands are present in the experimental data, the sample is reliably assigned to the appropriate group.

This study finds the genotype of an individual (in this case, the genotype related to sickle cell anemia in a prenatal human), from a measure of the amount of a genetic marker (in this case the mutation site in the hemoglobin gene shown in Table 1 on page 5082) designated as a zero or non-zero amount. The analysis of Geever et al. was repeated on a plurality of individuals.

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to practice the art claimed in these claims of the application thus resulting in the practice of the instantly claimed invention with a reasonable expectation of success because while Anderson teaches a general probabilistic clustering method to analyze expression of genes Geever teaches the advantages of genotyping SNPs in humans, and analyzing the genotype and allelic markers to indicate sickle cell anemia.

Claims 1 and 4 are rejected under 35 U.S.C. 103(a) as being unpatentable over Anderson in view of Geever as applied to claims 1-3, 5-7, 63, and 73 above, and further in view of Xue et al. [PGPUB 2003/0017487].

1. A method for determining a genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: assigning the measure of the amount of the allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group, wherein the individual is determined to have the genotype assigned to the group.

4. A method as claimed in claim 1, wherein the individual is a haploid organism.

While Anderson et al and Geever et al teach the method of claim 1, that is the use of finding markers (SNPs) to determine genotypes, they fail to teach such analyses for haploid organisms.

Xue et al. in Figure 2 does teach SNP analysis of haploid cells.

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to practice Anderson et al in view of Geever et al in further view of Xue et al. because Xue et al has the advantage of examining SNPs in haploid cells.

Claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75,and 76 are rejected under 35 U.S.C. 103(a) as being unpatentable over Anderson et al in view of Geever et al as applied to claims 1-3, 5-7, 63, and 73 above, and further in view of Krishna et al [IEEE Transactions of Systems, Man, and Cybernetics—Part B: Cybernetics, volume 29, June 1999, pages 433-439].

Claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75,and 76 state:

1. A method for determining a genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: assigning the measure of the amount of the allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group, wherein the individual is determined to have the genotype assigned to the group.

9. A method as claimed in claim 1, wherein the distance-based clustering process carries out at least one K-means algorithm.

10. A method as claimed in claim 9, wherein the at least one K-means algorithm is initiated by assigning a plurality of mean values evenly distributed between approximately 0 and approximately 1.

11. A method as claimed in claim 10, wherein 10 mean values are assigned.

12. A method as claimed in claim 9, wherein the at least one K-means algorithm determines a solution for a plurality of subsets of density centers.

13. A method as claimed in claim 12, wherein each subset comprises three density center values.

16. A method as claimed in claim 12, wherein the plurality of subsets of density center values comprises every combination of subsets of density center values.

17. A method as claimed in claim 9, comprising carrying out a first K-means algorithm and a second K-means algorithm.

18. A method as claimed in claim 17, comprising carrying out a first K-means algorithm and a second K-means algorithm, wherein the first K-means algorithm is initiated by assigning a plurality of mean values; and the second K-means algorithm determines a solution for a plurality of subsets of density centers obtained by the first K-means algorithm.

20. A method as claimed in claim 1, wherein the probability clustering process is initiated using a solution obtained by at least one K-means algorithm.

21. A method as claimed in claim 1, wherein a solution obtained by the probability clustering process and/or the distance-based clustering process yields a minimum maximum standard deviation for a distribution of the at least one measure of the amount of the allele.

22. A method as claimed in claim 1, comprising assigning the measure of the amount of the allele using both a probability clustering process and a distance-based clustering process.

25. A method as claimed in claim 1, wherein the property of the group is a characterizing property of the group and the genotype is assigned based on the characterizing property of the group falling within a one of a plurality of ranges of values of the measure of the amount of the allele, each of the plurality of ranges of values corresponding to a different genotype.

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the genotype assigned to the group.

65. A data processing apparatus as claimed in claim 63, wherein the computer code executing a distance-based clustering process carries out at least one K-means algorithm.

67. A data processing apparatus as claimed in claim 63, wherein the probability clustering process is initiated using a solution obtained by at least one K-means algorithm.

68. A data processing apparatus as claimed in claim 63, wherein a solution obtained by the probability clustering process and/or the distance-based clustering process yields a minimum maximum standard deviation for a distribution of the at least one measure of the amount of the allele.

69. A data processing apparatus as claimed in claim 67, wherein the computer code executes both a probability clustering process and a distance-based clustering process.

70. A data processing apparatus as claimed in claim 69, wherein the probability clustering process carries out an expectation maximization algorithm and the distance-based clustering process carries out at least one K-means algorithm.

73. A computer readable medium comprising computer readable code for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of an allele of the genetic marker in the individual, and for carrying out the processes of: assigning the measure of the amount of an allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group and determining the individual to have the genotype assigned to the group.

75. A computer readable medium as claimed in claim 73, wherein the distance-based clustering process comprises a K-means algorithm.

76. A computer readable medium as claimed in claim 73, wherein the measure of the amount of the allele is assigned using both a probability clustering process and a distance-based clustering process.

While Anderson et al in view of Geever et al teach the method of determining a genotype from a genetic marker, they do not teach analysis in terms of K-means genetic algorithms.

The article of Krishna et al, entitled, "Genetic K-Means Algorithm," states in its abstract:

In this paper, we propose a novel hybrid genetic algorithm (GA) that finds a globally optimal partition of a given data into a specified number of clusters. GA's used earlier in clustering employ either an expensive crossover operator to generate valid child chromosomes from parent chromosomes or a costly fitness function or both.   To circumvent these expensive operations, we hybridize GA with a classical gradient descent algorithm viz., K-means algorithm. Hence, the name genetic K-means algorithm (GKA). We define K-means operator, one-step of K-means algorithm, and use it in GKA as a search operator instead of crossover. We also define a biased mutation operator specific to clustering called distance-based-mutation. Using finite Markov chain theory, we prove that the GKA converges to the global optimum. It is observed in the simulations that GKA converges to the best known optimum corresponding to the given data in concurrence with the convergence result. It is also observed that GKA searches faster than some of the other evolutionary algorithms used for clustering.

Tables I and II on page 348 of Krishna et al illustrate the assignation of 10 means with the plurality of mean values distributed between "approximately" 0 and "approximately" 1. Equation 3 on page 434 illustrates a plurality or greater than three centroids or density centers.

The actual K-means algorithm is illustrates on the bottom of column 1 on page 436 of Krishna et al. It can be carried out multiple times until optimization to a solution as it is iterative.

Equations 4 and 5 of Krishna et al on column 2 of page 434 list within-cluster variations and total within cluster variations used to evaluate variances.

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to modify Anderson et al in view of Geever et al as applied to claims 1-3, 5-7, 63, and 73 above, and further in view of Krishna et al because Krishna et al has the advantage of employing a genetic K means cluster analysis for faster analysis of the data.

Claims 1, 8, 63-64, and 73-74 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Anderson et al in view of Geever et al as applied to claims 1-3, 5-7,

63, and 73 above, and further in view of Excoffier et al [Mol. Biol. Evol. Volume 12,

pages 921-927, 1995].

Claims 1, 8, 63-64, and 73-74 state:

1. A method for determining a genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: assigning the measure of the amount of the allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group, wherein the individual is determined to have the genotype assigned to the group.

8. A method as claimed in claim 1, wherein the probability clustering process carries out an expectation maximization algorithm.

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the genotype assigned to the group.

64. A data processing apparatus as claimed in claim 63, wherein the computer code executing a probability clustering process carries out an expectation maximization algorithm.

73. A computer readable medium comprising computer readable code for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of an allele of the genetic marker in the individual, and for carrying out the processes of: assigning the measure of the amount of an allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group and determining the individual to have the genotype assigned to the group.

74. A computer readable medium as claimed in claim 73, wherein the probability clustering process comprises an expectation maximization algorithm.

While Anderson et al in view of Geever et al teach the method of determining a genotype from a genetic marker, they do not teach analysis in terms of an expectation-maximization (EM) algorithm.

The article of Excoffier et al, entitled, "Maximum likelihood of molecular haplotype frequencies in a diploid population," states in the abstract, "Molecular techniques allow the survey of a large number of linked polymorphic loci in random samples from diploid populations. …we implement an expectation-maximization (EM) algorithm leading to maximum-likelihood estimates of molecular haplotype frequencies under the assumption of Hardy Weinberg proportions."

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to modify Anderson et al in view of Geever et al as applied to claims 1-3, 5-7, 63, and 73 above, and further in view of Excoffier et al because Excoffier et al use EM algorithms for increase power and efficiency in surveying chromosomes.

Claims 1, 25-53, 63, 65, 73, and 77 are rejected under 35 U.S.C. 103(a) as being unpatentable over Anderson et al in view of Geever et al in view of Krishna et al as applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above, and further in view of Montoya-Delgado et al [Genetics, volume 158, pages 875-883, June 2001] in view of Frey et al [Journal of Immunological Methods, 1998, volume 221, pages 35-41].

Claims 1 and 25-53, 63, 65, 73, and 77 state:

1. A method for determining a genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: assigning the measure of the amount of the allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group, wherein the individual is determined to have the genotype assigned to the group.

25. A method as claimed in claim 1, wherein the property of the group is a characterizing property of the group and the genotype is assigned based on the characterizing property of the group falling within a one of a plurality of ranges of values of the measure of the amount of the allele, each of the plurality of ranges of values corresponding to a different genotype.

26. A method as claimed in claim 1, wherein the method determines the genotype of a plurality of individuals using a plurality of respective measures of the amount of the allele of the same genetic marker in each of the individuals, and wherein each of the measures of the amount of the allele is assigned to a one of a plurality of potential groups.

27. A method as claimed in claim 26, further comprising; assessing a confidence of the determinations of the genotypes of the plurality of individuals based on a criterion for p-values corresponding to a particular confidence level.

28. A method as claimed in claim 27, wherein assessing the confidence includes carrying out at least a first and a second evaluation of the confidence in the determinations.

29. A method as claimed in claim 28, wherein the first evaluation comprises a chi-squared distribution to determine the confidence in the determinations.

30. A method as claimed in claim 29, wherein a chi-squared distribution p-value is calculated for each standard deviation of each group from a chi-squared distribution based on a pre-set maximum standard deviation cutoff and a number of degrees of freedom reflecting the number of genotypes in a given group, and the determination of genotypes is rejected if one of the chi-squared distribution p-values does not meet a criterion corresponding to a confidence level.

31. A method as claimed in claim 30, wherein the maximum standard deviation cutoff is set to a value of 0.05.

32. A method as claimed in claim 30, wherein the confidence level is 99.9%.

33. A method as claimed in claim 28, wherein the second evaluation comprises determining a likelihood of the assigned distributions conforming to a corresponding

Hardy-Weinberg equilibrium for the plurality of individuals.

34. A method as claimed in claim 33, wherein determining the likelihood includes: calculating a first sum of a first set of Bayesian factors for all possible permutations of the plurality of individuals over the plurality of potential groups; calculating a second sum of Bayesian factors from a lowest Bayesian factor in the first set to the Bayesian factor corresponding to the assigned distribution of the individuals between the groups; and determining a p-value from the quotient of the second sum and the first sum.

35. A method as claimed in claim 34, wherein the determination of genotypes is rejected if the p-value does not meet a criterion corresponding to a confidence level.

36. A method as claimed in claim 35, wherein the confidence level is 99.9%.

37. A method as claimed in claim 28, further comprising determining a ratio between a likelihood of a measure of the amount of an allele corresponding to the assigned group and a likelihood of the measure of the amount of an allele corresponding to the next best fit group.

38. A method as claimed in claim 26, wherein at least one solution from the probability clustering process and the distance-based clustering process includes validating a number of groups to which the measures of the amount of the allele are to be assigned.

39. A method as claimed in claim 38, wherein validating the number of groups includes assigning the measures of the amount of the allele to a first number of groups and using a property of the groups to determine whether the assignment to the first number of groups is reliable.

40. A method as claimed in claim 39, wherein the property of the groups is a mean or median value of each group.

41. A method as claimed in claim 40, wherein using a property of the groups includes determining whether the mean or median values of the groups are sufficiently dissimilar for the groups to constitute different groups.

42. A method as claimed in claim 41, wherein the mean or median values of the groups are considered to be sufficiently dissimilar if they differ by more than a cut off corresponding to a difference which minimizes a number of incorrect genotype assignments for two independent genotype assignments for the same individuals.

43. A method as claimed in claim 39, wherein if it is determined that the assignment to the first number of groups is reliable, then genotypes are assigned to the groups.

44. A method as claimed in claim 39, wherein if it is determined that the assignment to

the first number of groups is unreliable, then one or more of the probability clustering process and the distance-based clustering process is repeated to assign the measures of the amount of the allele to a second number of groups different from the first number of groups.

45. A method as claimed in claim 44, wherein the second number of groups is less than the first number of groups.

46. A method as claimed in claim 44, wherein if the measures are assigned to three groups, then the genotype is assigned depending on a ranking of the groups, and if the measures are assigned to less than three groups, then the genotype is assigned depending on a mean value of each group.

47. A method as claimed in claim 46, wherein the genotype is selected from the group consisting of: homozygous reference; homozygous alternate; and heterozygous.

48. A method as claimed in claim 43, wherein the K-means clustering process includes determining a representative value for each group and assigning a genotype to each group based on the respective representative values of each group.

49. A method as claimed in claim 48, wherein assigning the genotype includes determining whether the representative value of a group falls within a one of a plurality of ranges of values.

50. A method as claimed in claim 48 or 49, wherein the representative value is a mean or median of a group.

51. A method as claimed in claim 49, wherein there are three ranges of values and a first range corresponds to a homozygous reference, a second range corresponds to a heterozygous and a third range corresponds to a homozygous alternate.

52. A method as claimed in claim 49, wherein the plurality of ranges of values have been determined by calibrating the measure of the amount of the allele using a sufficiently large sample of individuals to allow groups corresponding to all the different genotypes to be unambiguously determined.

53. A method as claimed in claim 52, wherein at least one boundary of each range is the value at which adjacent corresponding groups intersect.

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of

the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the genotype assigned to the group.

65. A data processing apparatus as claimed in claim 63, wherein the computer code executing a distance-based clustering process carries out at least one K-means algorithm.

73. A computer readable medium comprising computer readable code for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of an allele of the genetic marker in the individual, and for carrying out the processes of: assigning the measure of the amount of an allele to a group using one or more of a probability clustering process and a distance-based clustering process; and assigning a genotype to the group based on a property of the group and determining the individual to have the genotype assigned to the group.

77. A computer readable medium as claimed in claim 73, further comprising computer readable code for determining the confidence of the determination of genotype of at least one individual.

The article of Anderson et al in view of Geever et al in view of Krishna et al teach the clustering methods (distance and K means), but do not teach the required Bayesian analyses or the statistics involving levels of confidence.

The article of Montoya-Delgado et al, entitled, "An unconditional exact text for the Hardy-Weinberg Equilibrium Law: Sample-space ordering using the Bayes factor," uses the Hardy Weinberg algorithm to look at ranges and ratios between the heterozygous and the two homozygous allele types (i.e. see the equations in column 1 of page 877 of Montoya-Delgado et al as well as Figure 1 of Montoya-Delgado et al). Figure 1 additionally illustrates the boundaries of the probabilities of the allele groups (i.e. the ranges for homozygous and heterozygous traits). Montoya Delgado et al also derives Bayesian groups as shown in the equations on page 878. Table 5 of Montoya-Delgado et al on page 881 teaches usage of p-values and chi-squared statistics.

The abstract of Montoya-Delgado et al states, "Much forensic inference based

upon DNA evidence is made assuming that the Hardy-Weinberg equilibrium (HWE) is

valid for the genetic loci being used. Several statistical tests to detect and measure

deviation from HWE have been derived, each having advantages and limitations....

Here we present an exact test for HWE in the biallelic case, based on the ratio of

weighted likelihoods under the null and alternative hypothesis, the Bayes factor. By

ordering the sample space using the Bayes factor, we also define a significance

(evidence) index, P values, using the weighted likelihood under the null hypothesis. We

compare it to the conditional exact test for the case of a sample size n = 10. Using the

idea under the method of chi square partition, the test is used sequentially to test

equilibrium in the multiple allele case and then applied to two short tandem repeat loci,

using a real Caucasian data bank, showing its usefulness."

However, Montoya-Delgado et al does not teach usage of confidence intervals,

or cutoff values.

The article of Frey et al, entitled, "A statistically defined endpoint titer

determination method for immunoassays," has an abstract which states:

Results of immunoassays for which no positive standards are available are often
expressed as endpoint titers. The endpoint titer is defined as the reciprocal of the
highest analyte dilution that gives a reading above the cutoff. Unfortunately, there is no
generally accepted rule for the determination of these cutoff values. In enzyme-linked
immunosorbent assays (ELISA) a value of two or three times the mean background or
negative control reading is sometimes used.... The procedure involves calculating the
upper prediction limit using the Student t-distribution. The mathematical formula which
defines the upper prediction limit is expressed as the standard deviation multiplied by a
factor which is based on the number of negative controls and the confidence level (1 −
alpha). Appropriate factors are provided for 2 to 30 negative controls and for
confidence levels ranging from 95% to 99.9%....

Table I on page 37 of Frey et al lists the required confidence levels of 99.5% and

99.9%.

It would have been obvious for someone of ordinary skill in the art at the time of

the instant invention to modify Anderson et al in view of Geever et al in view of Krishna

et al as applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above,

and further in view of Montoya-Delgado et al in view of Frey et al because Montoya-

Delgado et al teach the use of the required algorithms for better forensic inference and

Frey et al teaches the appropriate statistical analysis for better differentiation of the

results of immunoassays.

Claims 58-62, 63, 67, and 69-72 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Anderson et al in view of Geever et al in view of Krishna et al as

applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above, and

further in view of Excoffier et al.

Claims 58-62, 63, 67, and 69-72 state:

58. A computer implemented method for determining one or more genotypes of a
plurality of individuals at a SNP position using respective measures of a relative allele
amount for the SNP position for each individual, comprising: assigning the measures of
the relative allele amount to a group using one or more of an expectation maximization
process and a K-means process; assigning a genotype to each group identified by the
expectation maximization process and/or the K-means process to determine a genotype
of each person; and assessing a confidence of determination of the genotype.

59. A method as claimed in claim 58, wherein the expectation maximization process is
initiated using a K-Means algorithm.

60. A method as claimed in claim 58, wherein assessing the confidence includes using
a chi-squared distribution to evaluate a spread of at least one of said groups and
evaluating whether a distribution of the individuals between the groups conforms to a
corresponding Hardy-Weinberg equilibrium distribution.

61. A method as claimed in claim 58, wherein the expectation maximization process determines a number of groups to which to assign the measures and assigns the measures to less than three groups if it is determined that an assignment to three groups would be unreliable.

62. A method as claimed in claim 61, wherein if the measures are assigned to three groups, then the genotype is assigned depending on the rank of the groups, and if the measures are assigned to less than three groups, then the genotype is assigned depending on a mean value of each group.

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the genotype assigned to the group.

67. A data processing apparatus as claimed in claim 63, wherein the probability clustering process is initiated using a solution obtained by at least one K-means algorithm.

69. A data processing apparatus as claimed in claim 67, wherein the computer code executes both a probability clustering process and a distance-based clustering process.

70. A data processing apparatus as claimed in claim 69, wherein the probability clustering process carries out an expectation maximization algorithm and the distance-based clustering process carries out at least one K-means algorithm.

71. A data processing apparatus as claimed in claim 70, wherein the computer code assigns the measure of the amount of the allele by comparing a) a solution determined by the K-means algorithm and b) a solution determined by the expectation-maximization solution, wherein the measure of the amount of the allele is assigned to a group according to the solution yielding the minimum maximum standard deviation.

72. A data processing apparatus as claimed in claim 63, further comprising computer code for determining a confidence of the determination of genotype of at least one individual.

While Anderson et al in view of Geever et al in view of Krishna et al teach the method of determining a genotype from a genetic marker, they do not teach analysis in terms of an expectation-maximization (EM) algorithm.

The article of Excoffier et al, entitled, "Maximum likelihood of molecular haplotype frequencies in a diploid population," states in the abstract, "Molecular techniques allow the survey of a large number of linked polymorphic loci in random samples from diploid populations. ...we implement an expectation-maximization (EM) algorithm leading to maximum-likelihood estimates of molecular haplotype frequencies under the assumption of Hardy Weinberg proportions."

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to modify Anderson et al in view of Geever et al as in view of Krishna as applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above, and further in view of Excoffier et al because Excoffier et al use EM algorithms for increase power and efficiency in surveying chromosomes.

Claims 63 and 66 are rejected under 35 U.S.C. 103(a) as being unpatentable over Anderson et al in view of Geever et al in view of Krishna et al as applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above, and further in view of Babu et al [Pattern Recognition Letters, volume 14, 1993, pages 763-769].

63. A data processing apparatus for determining the genotype of at least one individual from a genetic marker using at least one measure of the amount of a given allele of the genetic marker in the individual, comprising: a data processor; a storage device holding computer readable code in communication with the data processor, the computer readable code including: computer code which assigns the measure of the amount of the allele to a group by executing one or more of a probability clustering process and a distance-based clustering process; and computer code which assigns a genotype to the group based on a property of the group and determines the individual to have the

genotype assigned to the group.

66. A data processing apparatus as claimed in claim 63, wherein the probability clustering process is initiated using a seed value.

While Anderson et al in view of Geever et al in view of Krishna et al teach the method of determining a genotype from a genetic marker, they do not teach K means seeding algorithms.

The article of Babu et al, entitled, "A near optimal initial seed selection in K-means algorithm using a genetic algorithm," states in its abstract, "The K-means algorithm for clustering is very much dependency on the initial seed values. We use a genetic algorithm to find a near optimal partitioning of the given data set by selecting proper initial seed values in the K-means algorithm..."

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to modify Anderson et al in view of Geever et al as in view of Krishna as applied to claims 1, 9-13, 16-18, 20-22, 25, 63, 65, 67-70, 73, 75, and 76 above, and further in view of Babu et al because Babu et al has a systematic means of generating seed values from which the clustering is very much dependent.

Claims 1, 26, 38-39, 44, and 54-57 are rejected under 35 U.S.C. 103(a) as being unpatentable over Anderson et al in view of Geever et al in view of Krishna et al in view of Montoya-Delgado et al in view of Frey et al as applied to claims 1, 25-53, 63, 65, 73, and 77 above, and further in view of Babu et al.

Claims 54-57 claim:

54. A method as claimed in claim 44, wherein the probability clustering process uses a first set of seed values to assign the first number of groups and a second set of seed

values to assign the second number of groups.

55. A method as claimed in claim 54, wherein the seed values are means for respective groups.

56. A method as claimed in claim 55, wherein the first set of seed values comprises approximately 0.2, 0.5 and 0.8, and the second set of seed values comprises approximately 0.3 and 0.7.

57. A method as claimed in claim 55, wherein a ratio of the members of the first set of seed values is approximately 1:1:1 and a ratio of the members of the second set of seed values is approximately 1:1.

While Anderson et al in view of Geever et al in view of Krishna et al tin view of

Montoya-Delgado et al in view of Frey et al teach the method of determining a genotype

from a genetic marker, they do not teach K means seeding algorithms.

The article of Babu et al, entitled, "A near optimal initial seed selection in K-

means algorithm using a genetic algorithm," states in its abstract, "The K-means

algorithm for clustering is very much dependency on the initial seed values. We use a

genetic algorithm to find a near optimal partitioning of the given data set by selecting

proper initial seed values in the K-means algorithm..."

The algorithms on page 765 illustrate the algorithms for multiple seedings.

It would have been obvious to someone of ordinary skill in the art at the time of

the instant invention to modify Anderson et al in view of Geever et al as in view of

Krishna as applied to claims 1, 25-53, 63, 65, 73, and 77above, and further in view of

Babu et al because Babu et al has a systematic means of generating seed values from

which the clustering is very much dependent. It would have been further obvious to

optimize the seeding process by using the specific parameters in claims 55-57 because

these parameters make the seeding process more efficient.

Conclusion

No claim is allowed

Papers related to this application may be submitted to Technical Center 1600 by facsimile transmission. Papers should be faxed to Technical Center 1600 via the central PTO Fax Center. The faxing of such pages must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 (December 28, 1993)(See 37 CFR § 1.6(d)). The Central PTO Fax Center Number is (571) 273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Russell Negin, Ph.D., whose telephone number is (571) 272-1083. The examiner can normally be reached on Monday-Friday from 7am to 4pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's Supervisor, Andrew Wang, Supervisory Patent Examiner, can be reached at (571) 272-0811.
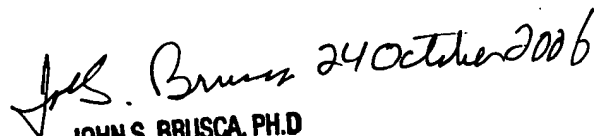
Any inquiry of a general nature or relating to the status of this application should be directed to Legal Instrument Examiner, Yolanda Chadwick, whose telephone number is (571) 272-0514.

Information regarding the status of the application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information on the PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

RSN
24 October 2006

24 October 2006

JOHN S. BRUSCA, PH.D
PRIMARY EXAMINER